Requirements Engineering

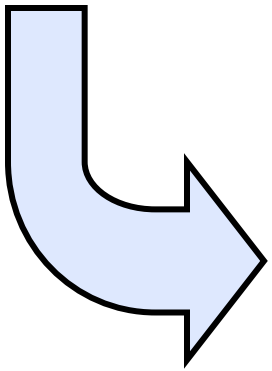
(Summer 2019)

Prof. Nan Niu (nan.niu@uc.edu)

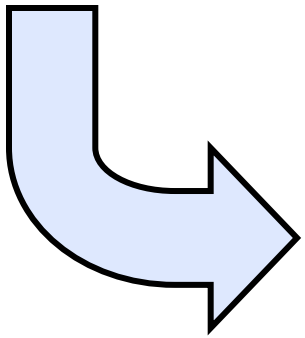
http://homepages.uc.edu/~niunn/courses

Today's Menu

Last Seminar:
NFRs
Release ASN2



This Seminar:
Automated Traceability
Explain ASN2

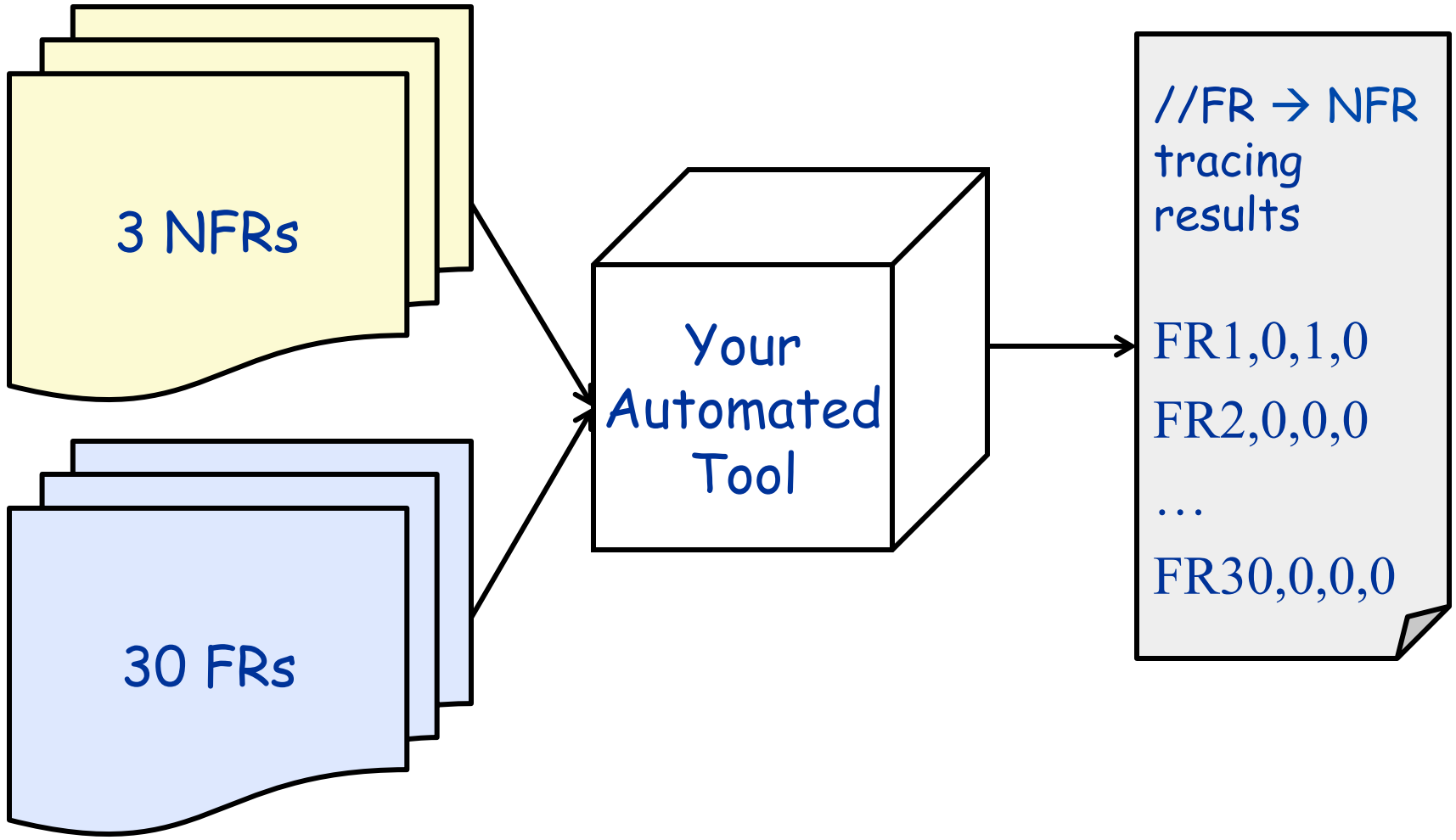


Next Seminar:
Carry out ASN3

Assignment 2

- Datasets available on the course website
  - http://homepages.uc.edu/~niunn/courses/
- Objectives
  - Automate the linking of FRs and NFRs
  - Sample linking algorithms
    - Keyword-based: |intersection| / |union|, Jaccard, …
    - Vector space model: tf-idf [Hayes-RE'03]
    - Probabilistic: NFR classifier [Cleland-Huang-RE'06]
    - …
- Due: in class, Wednesday (July 17)

ASN2: A conceptual picture



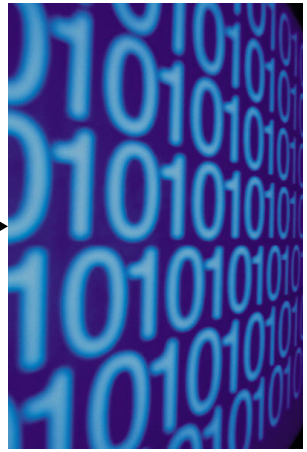
What's "req.s traceability"?

Yet another MIP awardee

[Gotel & Finkelstein, ICRE'94]

"***Requirements traceability*** *refers to the ability to describe and follow the life of a requirement, in both a forwards and backwards direction (i.e., from its origins, through its development and specification, to its subsequent deployment and use, and through all periods of on-going refinement and iteration in any of these phases)."*

FR3: The search radius shall be between 1 and 30 miles.



Why caring about "traceability"?

- Many standards consider it a quality indicator
  - IEEE STD-830-1998, "Guides to Software Requirements Specifications"
  - CMMI
  - U.S. Federal Aviation Administration (FAA)
  - …
- It is indispensable for carrying out many software engineering activities
  - Verification, e.g., whether code satisfies design
  - Validation, e.g., whether stakeholders' goals have been fulfilled
  - Change impact analysis, e.g., how much code will be affected if this requirement changes
  - System-level test coverage analysis
  - Risk assessment
  - …

Story about ChoicePoint

→ ChoicePoint

- Headquarters: Alpharetta (near Atlanta), Georgia, USA
- A data aggregation company
  - Combined personal data sourced from multiple public and private databases for sale to the government and the private sector
  - Maintained more than 17 billion records of individuals and businesses

→ Security breach

- In 2006, records on more than *163,000* consumers were acquired by identity thieves

→ Review by the US FCRA (Fair Credit Reporting Act)

- Revealed that ChoicePoint has developed the software products without proper controls mandated by the FCRA
- ChoicePoint was fined ***$15 million*** in civil penalties
- ChoicePoint must undergo biennial security audits for the next *20 years*

Source: *United States v. ChoicePoint, Inc., Case 1:06-CV-00198-JTC, (Northern District of Georgia), Feb 2006*

Story

- ChoicePoint
  - Headquarters: Alp
  - A data aggregatio
    - Combined personal databases for sale
    - Maintained more t
- Security breach
  - In 2006, records on more than *163,000* consumers were acquired by identity thieves
- Review by the US FCRA (Fair Credit Reporting Act)
  - Revealed that ChoicePoint has developed the software products without proper controls mandated by the FCRA
  - ChoicePoint was fined *$15 million* in civil penalties
  - ChoicePoint must undergo biennial security audits for the next *20 years*

Source: *United States v. ChoicePoint, Inc., Case 1:06-CV-00198-JTC, (Northern District of Georgia), Feb 2006*

Back to traceability: Can't we use spreadsheet?

Sample traceability matrix

Requirement Identifiers	Reqs Tested	REQ1 UC 1.1	REQ1 UC 1.2	REQ1 UC 1.3	REQ1 UC 2.1	REQ1 UC 2.2	REQ1 UC 2.3.1	REQ1 UC 2.3.2	REQ1 UC 2.3.3	REQ1 UC 2.4	REQ1 UC 3.1	REQ1 UC 3.2	REQ1 TECH 1.1	REQ1 TECH 1.2	REQ1 TECH 1.3
Test Cases	321	3	2	3	1	1	1	1	1	1	2	3	1	1	1
Tested Implicitly	77														
TC1.1.1	1	x													
TC1.1.2	2		x	x											
TC1.1.3	2	x											x		
TC1.1.4	1			x											
TC1.1.5	2	x												x	
TC1.1.6	1		x												
TC1.1.7	1			x											
TC1.2.1	2				x		x								
TC1.2.2	2					x		x							
TC1.2.3	2								x	x					
TC1.3.1	1										x				
TC1.3.2	1										x				
TC1.3.3	1											x			
TC1.3.4	1											x			
TC1.3.5	1											x			
etc															
TC5.6.2	1														x

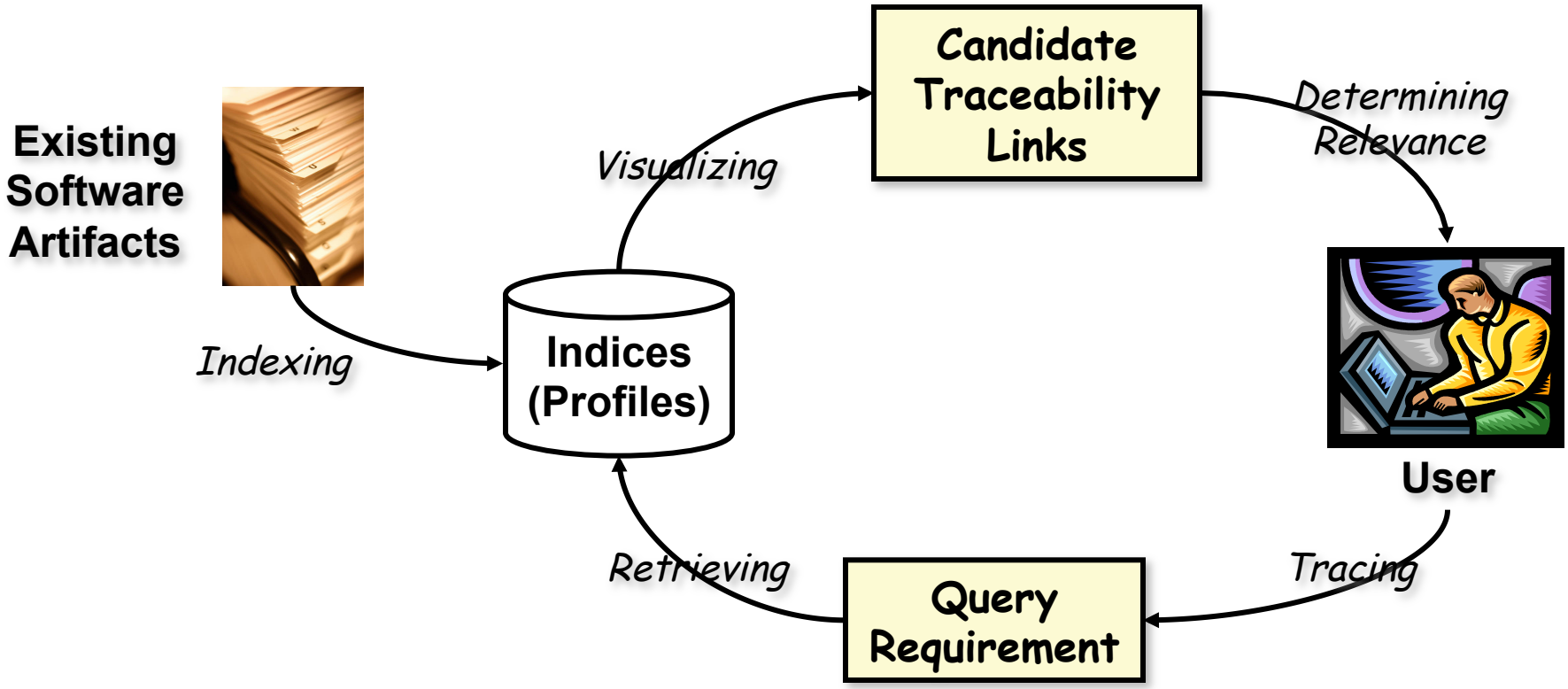
Spreadsheet (tracing manually in general) doesn't work

→ Tedious, time-consuming, & error-prone

- Scalability, e.g., hundreds of requirements & code files
- Evolving, i.e., keeping up with the changing software in a spreadsheet is not always a good use of your time

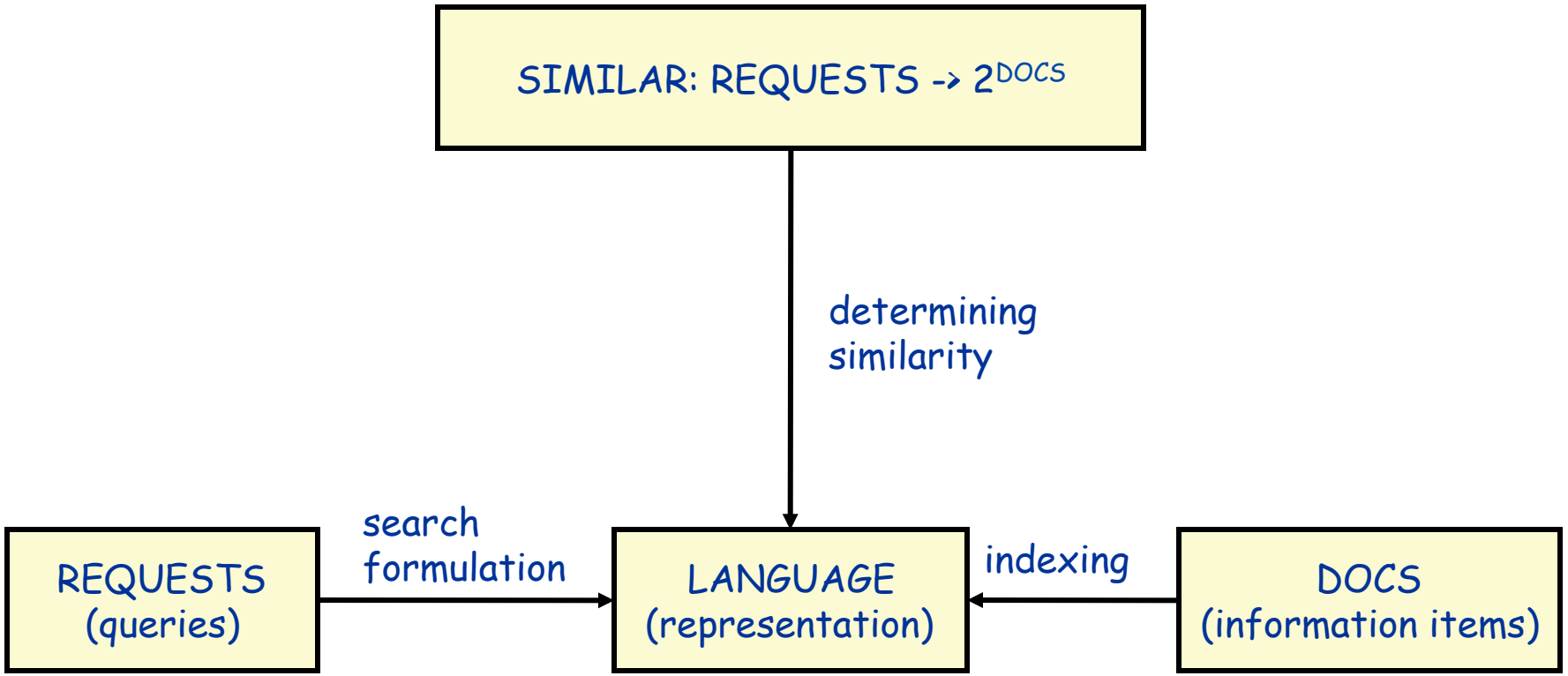
In practice, traceability is often dropped, or performed as needed (as opposed to systematically).

Automated Traceability (after-the-fact traceability)



Information Retrieval (IR)

hint: your ASN2 can be based on IR



Example

- Two requirements
  - r1 = "create and deactivate patients profile"
  - r2 = "patients create and edit profile"
- In this lecture, we introduce some basic retrieval methods: set-based, Jaccard.
- Assumption of IR-based ASN2 solution
  - the more textual similarity there is between the two requirements, the more likely one is linked with (traceable to) the other

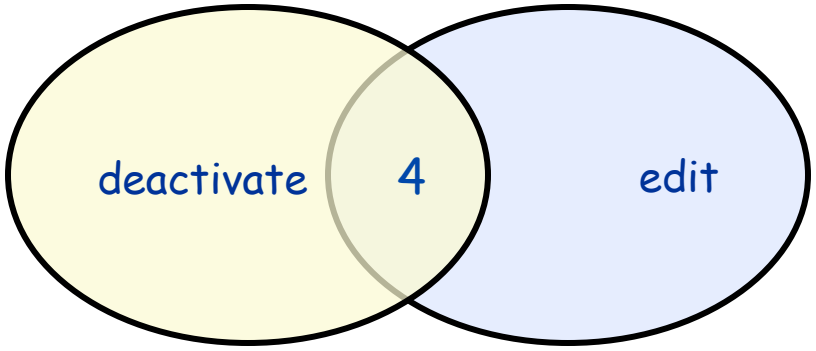
Similarity based on set overlapping

→ Basic formula

$$S(R1,R2) = \frac{2\,|R1 \cap R2|}{(|R1| + |R2|)}$$

↳r1 = "create and deactivate patients profile"

↳r2 = "patients create and edit profile"



→ Resulting similarity

↳ S(r1, r2) = (2x4) / (5+5) = 0.8

↳ Suppose the threshold is 0.5, then {r1, r2} would be regarded as traceable to each other

Similarity based on Jaccard index

→ Basic formula

The Jaccard similarity coefficient, J, is given as

$$J = \frac{M_{11}}{M_{01} + M_{10} + M_{11}}.$$

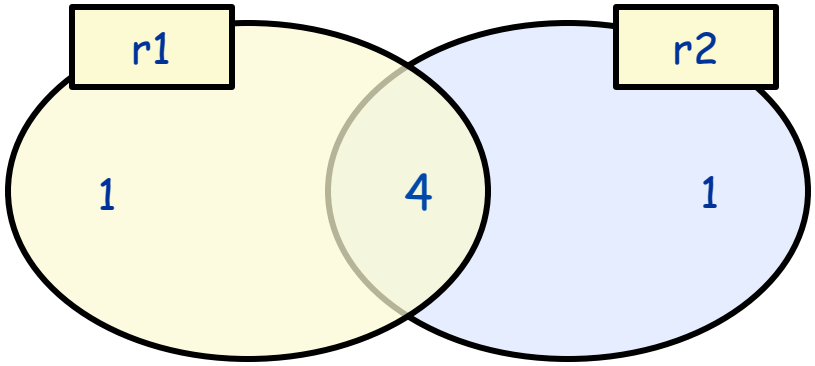
M_{11} represents the total number of attributes where A and B both have a value of 1.
M_{01} represents the total number of attributes where the attribute of A is 0 and the attribute of B is 1.
M_{10} represents the total number of attributes where the attribute of A is 1 and the attribute of B is 0.
M_{00} represents the total number of attributes where A and B both have a value of 0.

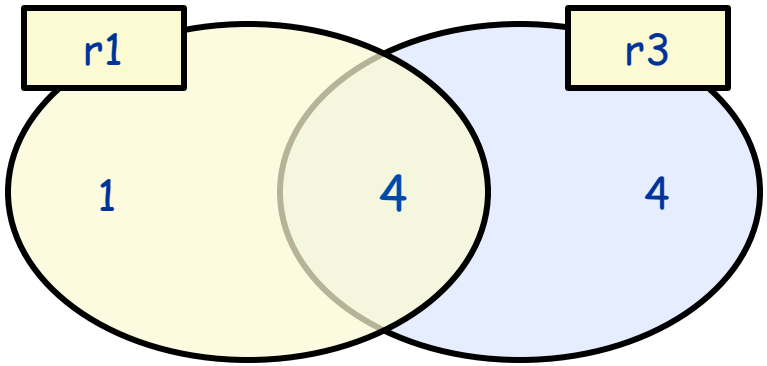
→ In our example

	create	and	deactivate	patients	profile	edit	a	including	photo
r1	1	1	1	1	1	0	0	0	0
r2	1	1	0	1	1	1	0	0	0
r3	1	1	0	1	1	1	1	1	1

Jaccard (cont'd)

- r1 = "create and deactivate patients profile"
- r2 = "patients create and edit profile"
- r3 = "patients create and edit profile including a photo"





→ Set-based similarity

- S(r1, r2) = (2x4) / (5+5) = 0.8
- S(r1, r3) = (2x4) / (5+8) = 0.62

→ Jaccard-based similarity

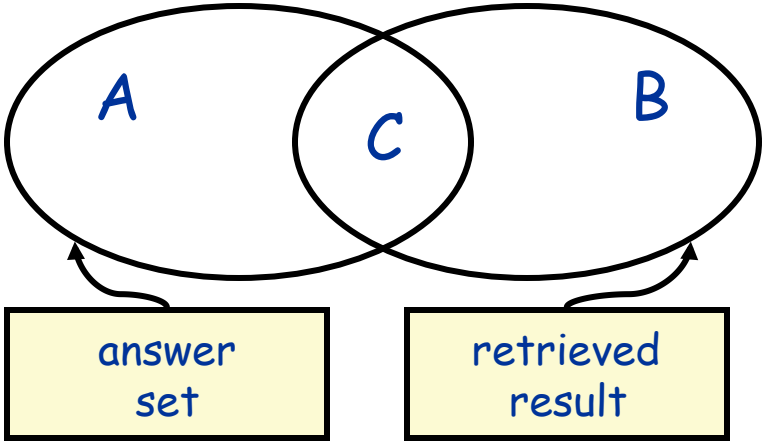
- S(r1, r2) = 4 / 6 = 0.67
- S(r1, r3) = 4 / 9 = 0.44

Evaluating Your ASN2 Solution

- The output of your ASN2 algorithm will be assessed via IR metrics
  - Recall, Precision, and F2

- Your ASN2 algorithm will be run twice on Wednesday (July 17)
  - 30 FRs and 3 NFRs
  - 36 FRs and 3 NFRs
  - That is, the 3 NFRs stay the same between the two runs, but 6 new FRs will be added to test your algorithm's performance with unseen data

IR Metrics



Precision (accuracy) = |C| / |B|
Recall (coverage) = |C| / |A|

$$\text{F-measure} = \frac{(1+\text{beta}^2) \times (P \times R)}{(\text{beta}^2 \times P + R)}$$

(F2-measure: beta=2; weights R twice as much as P)

Suppose |A|=10

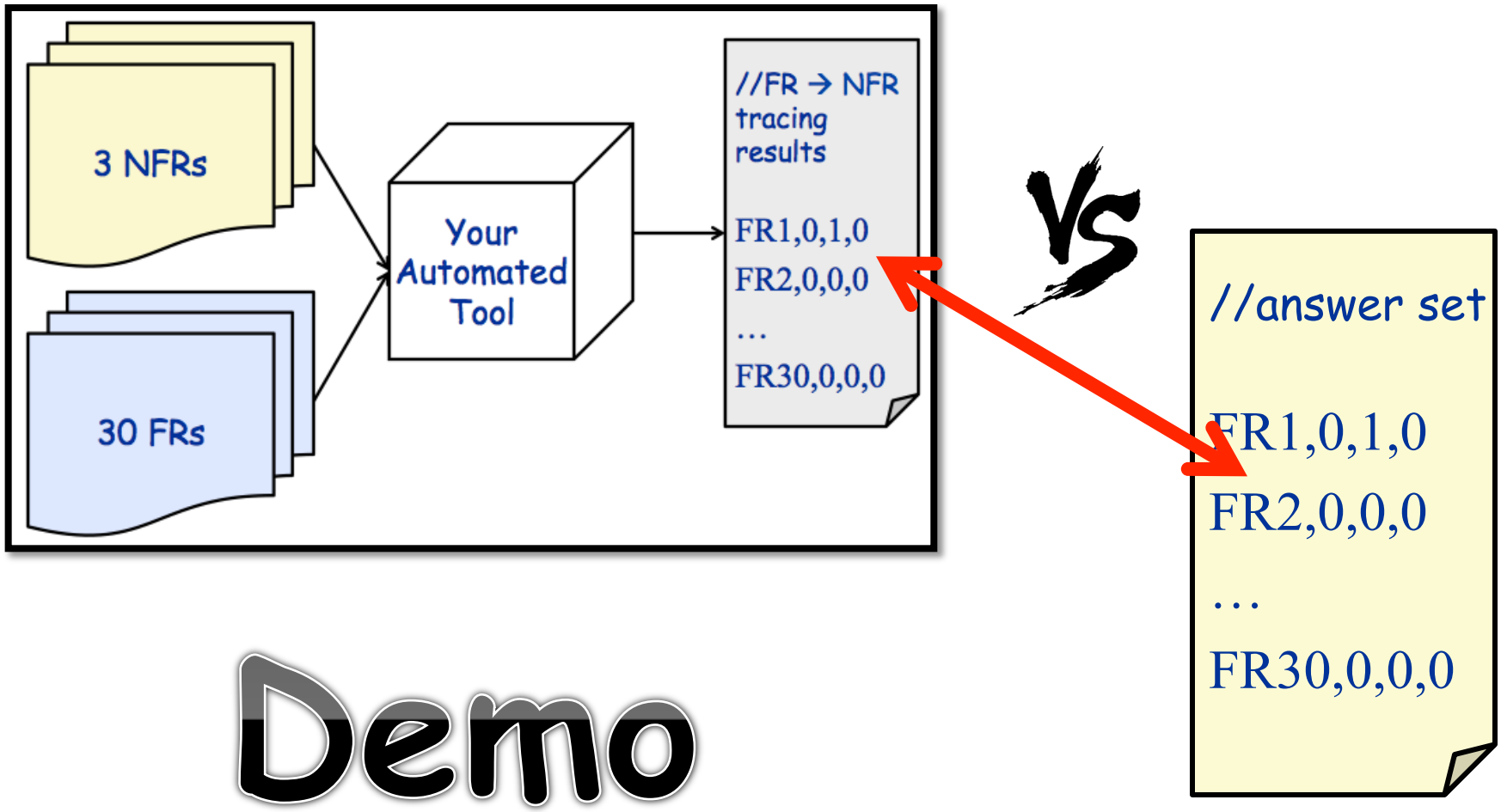
Result	Relevant
gr1	Yes
gr2	No
gr3	Yes
gr4	Yes
gr5	No

$\text{Precision}_{Google}$ = 3 / 5 = 60%
Recall_{Google} = 3 / 10 = 30%
$F2_{Google}$ = 0.33

Precision_{Bing} = 2 / 3 = 67%
Recall_{Bing} = 2 / 10 = 20%
$F2_{Bing}$ = 0.23

Result	Relevant
br1	Yes
br2	Yes
br3	No

ASN2 Performance Evaluation



Demo

Evaluating Your ASN2 Solution

→ The output of your ASN2 algorithm will be assessed via IR metrics

  ↳ Recall, Precision, and F2

→ Your ASN2 algorithm will be run twice on Wednesday (July 17)

  ↳ 30 FRs and 3 NFRs

  ↳ 36 FRs and 3 NFRs

  ↳ That is, the 3 NFRs stay the same between the two runs, but 6 new FRs will be added to test your algorithm's performance with unseen data